基于 STM 分析旅行者对不同档次酒店的偏好差异

杨海霞　吴维芳　孙含林
(武汉大学经济与管理学院　武汉　430072)

摘要: 【目的】探测 Web2.0 时代下，不同类型旅行者对不同档次酒店提供的产品或服务的偏好差异。【方法】将酒店划分为豪华型酒店和经济型酒店，将旅行者划分为 5 种旅行类型，分析不同旅行者对豪华型酒店和经济型酒店的评分模式，并借助结构主题模型(Structure Topic Model, STM)对酒店在线评论文本进行细致分析，挖掘在线评论话题，分析各个旅行类型的旅行者对不同档次酒店提供的服务偏好差异。【结果】实验结果表明: 5 种旅行类型下，旅行者对豪华型酒店的平均评分均高于对经济型酒店的平均评分; 各类旅行者，对不同档次酒店所提供的产品或服务存在偏好差异。【局限】实验数据不够充分; 忽略了诸如性别、年龄等因素对在线评论的数值评分和文本内容的影响。【结论】分析不同类型旅行者对不同档次酒店的偏好差异，有助于酒店管理者制定服务供应策略，有助于消费者制定购买决策。

关键词: 在线评论　评分模式　酒店档次　旅行方式　STM　评论话题
分类号: F59　G350

1　引　言

随着 Web2.0 时代的快速发展，越来越多用户在网络平台上发表观点。在线评论网站是一种非常流行的 Web2.0 应用[1]，允许用户以几乎为零的成本在网络上发布他们对某个产品或服务的购买体验[2]，形成用户生成内容(User Generated Content, UGC)。研究表明，UGC 不仅能帮助潜在消费者做出购买决策，还能帮助相关管理者提高其产品或服务的质量，获得更多收益[3-5]。

相关研究表明，不同旅行类型的旅行者对酒店服务的偏好不同[6-9]; 同时，不同档次酒店的定位不同，目标市场不同，因而消费者对其期望不同，产生的入住体验也不同[10]。然而，以往学者对酒店在线评论进行文本分析时，多集中于话题提取[11-13]，鲜有学者考虑消费者旅行类型和酒店档次，分析消费者对酒店服务的偏好差异。因此，本文按旅行者的出行方式[14-15]，将旅行者划分为“商务旅行”、“夫妻旅行”、“家庭旅行”、“朋友旅行”、“独自旅行”; 按照酒店星级，将酒店分为豪华型酒店和经济型酒店[16]，研究各类旅行者对不同档次酒店的评分差异以及服务偏好的差异。

本文对酒店在线评论的数值评分进行基本描述性统计分析，借助结构主题模型(Structural Topic Models, STM)[17-19]对评论文本内容进行话题提取，并加入“酒店档次”作为主题强度和主题内容的协变量，分别分析各类旅行者对不同档次酒店的服务偏好差异。

2　文献综述

酒店在线评论是旅行者在酒店网站或第三方评论网站上生成的对入住酒店的体验评价[2-5]。在线评论能够吸引潜在消费者访问，增加访问时间，并提升卖家与买家的关系质量。现有研究对在线评论的讨论多集于在评论数量、在线评论得分、在线评论效价、在线评论极性以及在线评分差异性等对产品销量或购买意愿的影响，且这些变量多以评分为背景[20-21]，很少涉

通讯作者: 杨海霞，ORCID: 0000-0002-4418-9488，E-mail: haixiayang@whu.edu.cn。

及评论的文本内容。

旅行者在 TripAdvisor.com 撰写酒店评论时，通常会被问到“这次旅行属于哪种类型(What sort of trip was this)”，旅行者需要在 5 种旅行类型(商务、夫妻、家庭、朋友和独自旅行)中，选择一种属于他的旅行类型[14-15]。以往研究表明，不同旅行者的旅行目的和期望的不同[22-23],会导致其对酒店提供的设施或服务的偏好亦不同，如位置(Location)、卫生(Cleanliness)、性价比(Value for Money)等。Rivers 等[6]的研究结果表明，商务出行的旅行者非常关注酒店位置的便利性及网络的可获得性，较少关注酒店的价格；Lee 等[7]的研究结果表明，夫妻出行的旅行者对酒店的满意程度容易受到目的地气候变化及浪漫程度的影响；Lai 等[8]的研究结果表明，与独自出行的旅行者相比，与家庭或者朋友一起出行的旅行者更强调安全。

国内在游客旅游偏好方面的研究较于国外起步晚。1987 年，保继刚[24]分析了旅游者偏好、旅游动机、旅游者行为与旅游地空间相互作用的关系。2006 年，杨瑞[25]运用回归分析方法，分析西安市大学生游客群体的出行偏好、需求特征和消费特征，并比较了大学生和普通游客的旅游出行偏好行为。2010 年，岳东菊等[26]利用调查问卷方法分析西安市国内旅游者的旅游偏好行为。2011 年，雷俐丽[27]对大连市游客的旅游动机、旅游体验和旅游产品进行了偏好分析，并对比研究了不同出行类型的游客的旅游行为特征。

然而，目前缺少酒店档次与消费者偏好相关的研究成果，虽然顾秀玲[28]在对环太湖度假酒店 4 类顾客消费者行为进行的实证研究中，考虑了游客对不同星级酒店的评价，但并未涉及评论文本内容。然而对酒店在线评论文本进行话题发现分析[29-32]，能有效发现不同类型旅行者对酒店偏好的差异。因此，本文借助无监督的主题模型 STM 分析各类旅行者对不同酒店档次的偏好差异。

STM 作为主题模型[33-34]的成员之一，假定文档是若干主题的混合，主题为若干词的混合。笔者借助 STM 模型分析不同旅行类型下评论者对酒店的文本评论差异。此外，STM 允许研究者加入其需要控制的变量作为协变量，对研究对象进行比较分析。本文主要加入“酒店档次”作为协变量，对主题强度和主题内容进行分析，探测各个旅行类型下的旅行者对不同档次酒店提供的产品或服务的偏好差异。

3 STM 模型

与其他主题模型[33-34]一样，STM 是生成模型。这意味着，STM为每个文档都定义了一个数据生成过程，借助数据找到最适合该模型的参数。在STM生成模型中，文档$(D_1,D_2,\cdots,D_n)$、主题$(T_1,T_2,\cdots,T_n)$、及主题–词$(w_1,w_2,\cdots,w_n)$分布生成与元数据X_d(d表示文档)相联系的文档。STM与LDA(Latent Dirichilet Allocation)[34]等主题模型一样，主题被定义为若干词的混合，且每个词都以一定的概率属于某个主题；文档是若干主题的混合，即一条评论可以由多个话题组成。因此，每条评论的所有话题的概率之和为1，一个词在所有话题中的概率之和为1。

在文档的生成过程中，表明主题强度(Topic Prevalence)和主题内容(Topic Content)可以表示为文档元数据的函数。主题强度表明一条评论中有多少内容与一个话题相关，话题内容用话题中的词表示。因此，主题强度协变量表示主题强度，主题内容协变量解释主题内容。需要注意的是，STM既可以只包含主题强度协变量或主题内容协变量中的一个，也可以不包含任何协变量。

在主题数为k的STM模型中，每条评论(文档d)的生成过程可以被描述如下：

(1) 从基于文档协变量X_d的广义线性模型中，生成文档–主题$\vec{\theta}_d \mid X_d\gamma, \Sigma \sim \text{Logistic}(\mu = X_d\gamma, \Sigma)$；

(2) 用基准词分布(m)，主题偏差(K_k)，协变量偏差(K_g)及其交互向偏差$K_{i=(kgd)}$，生成能表示每个主题(k)的词分布，$\beta_{d,k} \propto \exp(m + K_k + K_{gd} + K_{i=(kgd)})$；

(3) 对于文档中的每个词，$n(n \in 1,\cdots,N_d)$：

①基于文档特定的主题分布，将词分配给主题，$z_{d,n} \mid \vec{\theta}_d \sim \text{Multinomial}(\vec{\theta})$；

②从被选定的主题中生成一个词，$w_{d,n} \mid z_{d,n}, \beta_{d,k=z} \sim \text{Multinomial}(\beta_{d,k=z})$。

在本文中，笔者将“酒店档次(豪华型酒店和经济型酒店)”作为协变量，用结构主题模型 STM 估计酒店在线评论的文档–主题(Document-Topic)概率和主题–词(Word-Topic)概率，分析各类旅行者对不同档次酒店提供的服务的偏好差异。同时，为拟合模型，笔者借

助半–倒塌变分EM算法(Semi-Collapsed Variational Expectation-Maximization Algorithm)[17-19]，根据收敛度 (Convergence) 估计模型的参数，进而得到每条评论的主题分布θ_d。

4　实验与结果分析

4.1　数据来源

TripAdvisor.com 是全球知名的旅游评论网站，其每月独立访问人数已经达到 3.5 亿。在学术研究领域中，其是最受关注的酒店评论网站之一，近年来，许多研究者选择其中的相关数据作为数据集进行研究[2,35-37]。因而，本文选取 TripAdvisor.com 的酒店在线评论进行实验分析，具有一定的实践意义。

笔者于 2013 年 10 月采集 2012 年 1 月 1 日至 2013 年 9 月 30 日拉斯维加斯酒店的相关数据。对每一条酒店评论，收集评论的 ID、总体评分(Overall Rating)、评论文本(Review Text)、旅行类型(Profile)、住宿时间(Trip Time)、被评论的酒店星级(Star Class)和酒店 ID 等，其中旅行类型包括：商务(Business)、夫妻(Couple)、家庭(Family)、朋友(Friend)和独自旅行(Solo)。为分析评论文本的具体内容，删除非英文文本评论，最后得到 101 846 条评论。所采集的酒店数据中，酒店星级分为：1.5、2、2.5、3、3.5、4、4.5、5。笔者按照酒店星级将酒店大致分成两个档次：星级≥4(4、4.5、5)的酒店为豪华型酒店；星级＜4(1.5、2、2.5、3、3.5)的酒店为经济型酒店[16]。最后得到豪华型酒店评论 68 317 条，经济型酒店评论 33 529 条。

4.2　描述性统计分析

分别对每类旅行者对不同档次酒店的评分进行描述性统计分析，如表 1 所示：

表 1　酒店在线评论评分的描述性统计

旅行类型 (Profile)	经济型酒店			豪华型酒店			合计		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
商务(Business)	4 416	3.53	1.20	10 437	4.06	1.10	14 853	3.90	1.16
夫妻(Couple)	11 963	3.77	1.11	30 094	4.31	0.95	42 057	4.16	1.03
家庭(Family)	8 414	3.80	1.17	12 385	4.26	1.01	20 799	4.07	1.10
朋友(Friends)	6 397	3.69	1.13	12 902	4.21	1.01	19 299	4.04	1.08
独自旅行(Sole)	2 339	3.78	1.12	2 499	4.28	0.98	4 838	4.04	1.08

(注：N 表示评论的数量，Mean 表示评分的均值，SD 表示评分的标准差。)

分析可知，豪华型酒店的在线评论总数远高于经济型酒店的在线评论总数，且豪华型酒店比经济型酒店获得更高的平均评分。此外，不同旅行类型的旅行者给出的酒店评分有差异：夫妻旅行者的总体平均评分最高；商务旅行者的总体平均评分最低[2]。考虑酒店的星级，发现无论是豪华型酒店还是经济型酒店，商务旅行者给出的平均评分均最低；而对于豪华型酒店，夫妻旅行者给出的平均评分最高；对于经济型酒店，家庭旅行者给出的平均评分最高。

酒店在线评论的平均得分，会影响潜在消费者对酒店购买决策[2-5]。然而，提升酒店的平均评分，需要关注旅行者选择该类酒店的原因及其所关心的酒店设施或服务，从而制定相应的产品策略，满足其实际需求。

4.3　旅行者对不同档次酒店的偏好差异分析

对酒店在线评论文本进行分析，挖掘在线评论话题，探测各类旅行者对不同档次酒店的服务偏好差异。

(1) STM 建模

为分析同一类型旅行者对不同档次酒店的服务偏好差异，笔者对文本分析数据进行配对：对同一类旅行者，随机选择 4 000 条酒店的在线评论(豪华型和经济型各 2 000 条)。借助 R 软件中的 stm 包[17]对评论文本进行预处理，主要包括：将所有单词转化为小写、去除停顿词、移除数字、移除标点符号、词干化处理。

在构建主题模型前，需要确定模型的主题数。这实际上也是模型选择问题，主题数太少的模型会产生内容宽泛的主题；主题数太多的模型会产生许多较小的子主题，难以辨识。Roberts 等[18]认为主题的质量和解释度主要应满足两个准则：

①主题内部有较高的一致性，如讨论“网络(Internet)”的主题词中，高频词应包括 Internet, WiFi, Fee 等；

②主题之间的高频词有一定的差异性，如主题“员工服务(Staff)”中的高频词 Staff 不应是主题“位置(Location)”中的高频词。

笔者设定主题数目为 5-15，通过比较主题质量，发现主题数目为 10 最符合本文的实验模型。因此，笔者将“酒店档次”作为 STM 的协变量，分别分析各类旅行者对豪华型酒店和经济型酒店的偏好差异。

(2) 偏好差异分析

在主题模型中，每条评论都可以表示为若干话题的概率混合。比如，依据商务旅行者对豪华型酒店的评论，可以计算出旅行者对每个话题的关注比例 θ_A；依据商务旅行对经济型酒店的评论，可以计算旅行者对每个话题的关注比例 θ_C。对于每一种旅行类型，使用 θ_A-θ_C，即可得到同一类旅行者对不同档次酒店的服务偏好差异。笔者定义：如两者相减的绝对值大于 0.03，则差异明显。

①商务旅行者

依图 1 可知，相比经济型酒店，旅行者在选择豪华型酒店时还关注景观(Topic8)；而相比豪华型酒店，旅行者选择经济型酒店时，会比较关注卫生(Topic3)和交通(Topic4)。相关研究表明，商务旅行者选择经济型酒店主要有两种可能：一是此类商务旅行者需自行承担住宿费用，经济能力有限，其关注酒店的性价比；二是此类商务旅行者因酒店位置的便利程度不得不选择某个经济型酒店[38]。

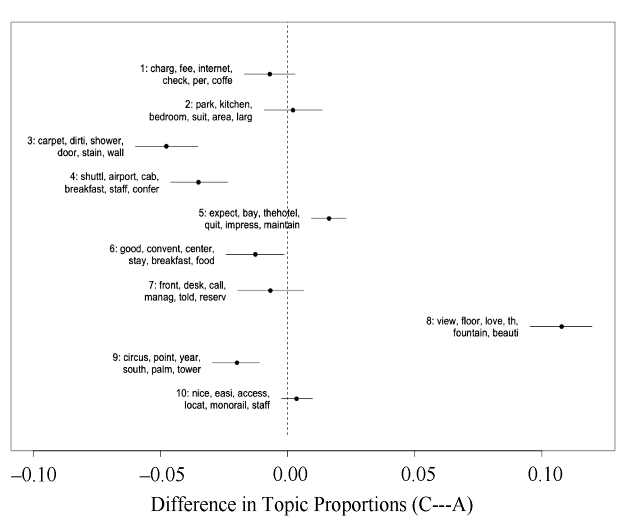


图 1　商务旅行者对豪华型酒店和经济型酒店的偏好差异

②夫妻旅行者

依图 2 可知，相比经济型酒店，夫妻旅行者在选择豪华型酒店时更多关注酒店的员工服务(Topic6)；而相比豪华型酒店，夫妻旅行者选择经济型酒店时，会比较关注酒店的卫生环境(Topic2)。相关研究表明，夫妻出行注重心情和浪漫度，其满意度容易受到其感知的浪漫度的影响[7]。对于高档型酒店，其提供的硬件设施服务相对比较完善，因而其软件服务(如员工服务)的提升对夫妻出行的满意度至关重要；而对于选择经济型酒店的夫妻旅行者，考虑到酒店的经济性，会较多关注酒店的卫生环境。

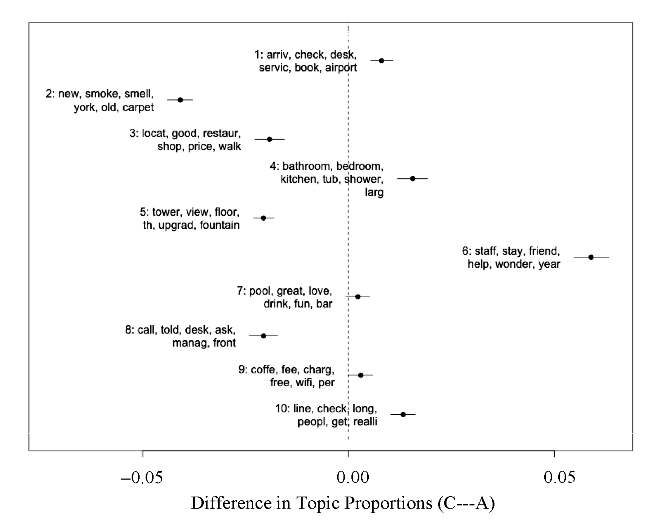


图 2　夫妻旅行者对豪华型酒店和经济型酒店的偏好差异

③家庭旅行者

依图 3 可知，相比经济型酒店，家庭旅行者在选择豪华型酒店时更多关注酒店的景色(Topic8)；而相比豪华型酒店，家庭旅行者选择经济型酒店时，会比较关注酒店的交通便利性(Topic6)，这是因为家庭旅行者通常都携带小孩或者老人[8]，即使经济能力欠佳，接受较低档次的酒店，也要注重交通的便利性。

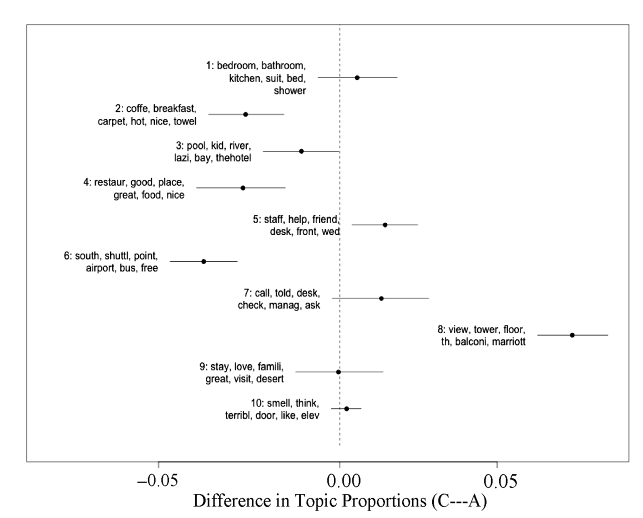


图 3　家庭旅行者对豪华型酒店和经济型酒店的偏好差异

④朋友旅行者

依图 4 可知，相比经济型酒店，朋友旅行者在选择豪华型酒店时更多关注酒店的氛围(Topic6)；而相比豪华型酒店，朋友旅行者选择经济型酒店时，会比较关注酒店的卧室特色(Topic7)。

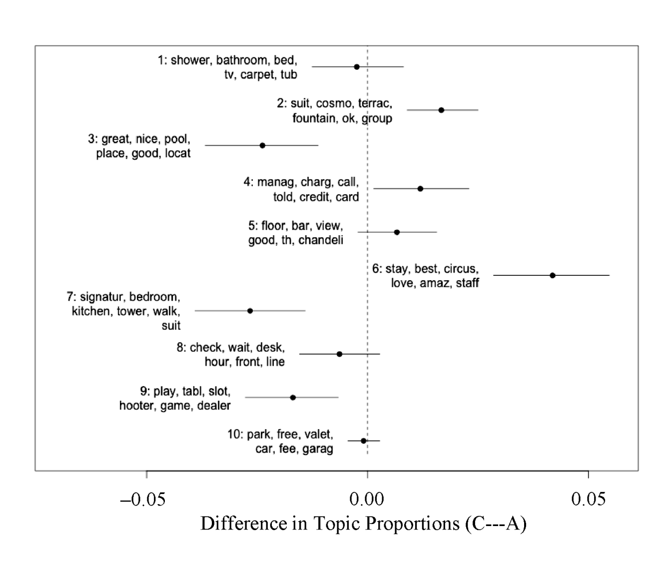


图 4　朋友旅行者对豪华型酒店和经济型酒店的偏好差异

⑤独自出行的旅行者

依图 5 可知，相比经济型酒店，独自出行的旅行者在选择豪华型酒店时更多关注酒店的整体感觉(Topic1)，包括位置、员工、卫生等。

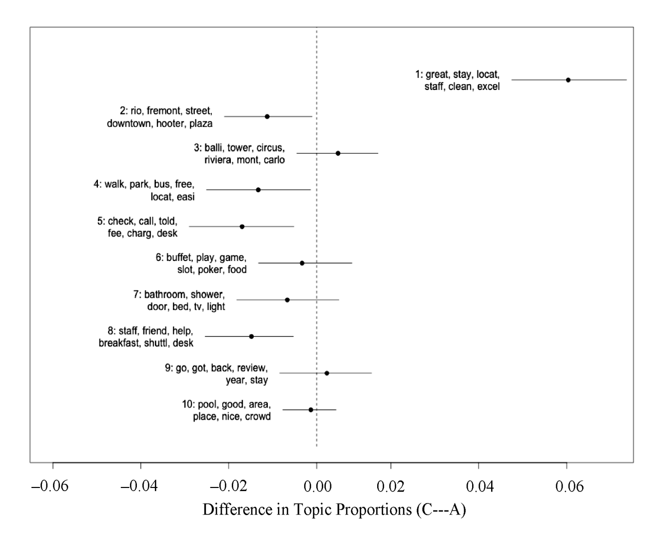


图 5　独自出行的旅行者对豪华型酒店和经济型酒店的偏好差异

5　结　语

5.1　局限性和未来的研究方向

笔者在现有研究的基础上，实现了一定的突破和改进，虽然存在一定的局限性，但这也为未来的研究提供了方向：

(1) 数据集仅来源于一个评论网站(TripAdvisor.com)中的一个城市(拉斯维加斯)，导致实验结果可能不具有充分的代表性，未来可以考虑从多个在线评论网站中采集数据，并包含多个城市的酒店评论数据，从地域方面对酒店在线评论进行文本分析。

(2) 笔者在 STM 模型中仅加入“酒店档次”作为协变量，仅考虑了旅行者对不同档次酒店的偏好差异，忽视了其他可能影响酒店服务偏好的因素(如评论者的性别、年龄等)。未来研究中，可以考虑加入更多协变量，探测不同旅行者对酒店服务的偏好差异。

5.2　管理启示

酒店档次对酒店评论评分有影响，豪华型酒店的总体平均评分高于经济型酒店的总体平均评分。豪华型酒店的设施齐全，提供的硬件设施服务也十分完善。因而对于豪华型酒店，若想提高酒店的总体评分，应加强员工的服务意识，满足顾客的心理需要。而经济型酒店为了保持较低的经营成本，无法提供更多的酒店设施，但其亦可以在服务态度上进行改善，提升酒店的员工服务质量，尤其是前台的入住手续办理服务和退房手续办理服务，顾客对酒店的第一印象和最后印象都会对评分产生重要影响。此外，经济型酒店还应格外加强对酒店的卫生管理，降低获得低分评论的概率。

不同旅行类型下的旅行者对不同档次酒店的偏差异也不同。商务旅行者评分较低且最为严格，夫妻旅行者评分较高且最为仁慈。对于豪华型酒店，酒店管理者可根据入住者的旅行类型，适当改变酒店服务供应策略，如为商务旅行者安排网络信号更好的房间，满足其商务办公需求；为夫妻旅行者安排视野更开阔的房间，满足其欣赏风景的需求。同理，对于经济型酒店，管理者也可根据旅行者类型，为商务旅行者提供网络更好的房间，为家庭旅行者提供更多的食物选择。同时，可以考虑为顾客提供免费的或低收费的接机服务，降低顾客寻找酒店地理位置的成本。

参考文献：

[1]　Buhalis D, Law R. Progress in Information Technology and Tourism Management: 20 Years on and 10 Years After the Internet—The State of eTourism Research [J]. Tourism Management, 2008, 29(4): 609-623.

[2]　Banerjee S, Chua A Y K. In Search of Patterns Among Travellers' Hotel Ratings in TripAdvisor [J]. Tourism Management, 2016, 53: 125-131.

[3] Brown J, Broderick A J, Lee N. Word of Mouth Communication Within Online Communities: Conceptualizing the Online Social Network [J]. Journal of Interactive Marketing, 2007, 21(3): 2-20.

[4] Goldenberg J, Libai B, Muller E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth [J]. Marketing Letters, 2001, 12(3): 211-223.

[5] Ye Q, Law R, Gu B. The Impact of Online User Reviews on Hotel Room Sales [J]. International Journal of Hospitality Management, 2009, 28(1): 180-182.

[6] Rivers M J, Toh R S, Alaoui M. Frequent-stayer Programs: the Demographic, Behavioral, and Attitudinal Characteristics of Hotel Steady Sleepers[J]. Journal of Travel Research, 1991, 30(2): 41-45.

[7] Lee C F, Huang H I, Chen W C. The Determinants of Honeymoon Destination Choice—The Case of Taiwan [J]. Journal of Travel & Tourism Marketing, 2010, 27(7): 676-693.

[8] Lai L, Graefe A R. Identifying Market Potential and Destination Choice Factors of Taiwanese Overseas Travelers [J]. Journal of Hospitality & Leisure Marketing, 1999, 6(4): 45-65.

[9] Poston R S. Using and Fixing Biased Rating Schemes [J]. Communications of the ACM, 2008, 51(9): 105-109.

[10] Munar A M, Jacobsen J K S. Motivations for Sharing Tourism Experiences Through Social Media [J]. Tourism Management, 2014, 43: 46-54.

[11] Mukherjee A, Liu B. Aspect Extraction Through Semi-supervised Modeling [C]. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. 2012: 339-348.

[12] Wang H, Lu Y, Zhai C. Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach [C]. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA. 2010: 783-792.

[13] Zhao Y, Dong S, Yang J. Effect Research of Aspects Extraction for Chinese Hotel Reviews Based on Machine Learning Method [J]. International Journal of Smart Home, 2015, 9(3): 23-34.

[14] Dolnicar S. Business Travellers' Hotel Expectations and Disappointments: A Different Perspective to Hotel Attribute Importance Investigation [J]. Asia Pacific Journal of Tourism Research, 2002, 7(1): 29-35.

[15] O'Connor P. User-generated Content and Travel: A Case Study on Tripadvisor.com [A]. // Information and Communication Technologies in Tourism 2008[M]. Springer Vienna, 2008.

[16] 高宝俊, 孙含琳, 王寒凝. 在线评论对酒店订满率的影响研究[J]. 旅游学刊, 2016, 31(4): 109-117. (Gao Baojun, Sun Hanlin, Wang Hanning. Influence of Online Reviews on Hotels' Full-occupancy Rates[J]. Tourism Tribune, 2016, 31(4): 109-117.)

[17] Roberts M E, Stewart B M, Tingley D. STM: R Package for Structural Topic Models [J]. General Information, 2014, 57(1): 445-460.

[18] Roberts M E, Stewart B M, Tingley D, et al. Structural Topic Models for Open-Ended Survey Responses [J]. American Journal of Political Science, 2014, 58(4): 1064-1082.

[19] Lucas C, Nielsen R A, Roberts M E, et al. Computer-Assisted Text Analysis for Comparative Politics [J]. Political Analysis, 2015, 23(2): 254-277.

[20] Liu Y. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue [J]. Journal of Marketing, 2006, 70(3): 74-89.

[21] Sun M. How Does the Variance of Product Ratings Matter? [J]. Management Science, 2012, 58(4): 696-707.

[22] Ariffin A A M, Maghzi A. A Preliminary Study on Customer Expectations of Hotel Hospitality: Influences of Personal and Hotel Factors [J]. International Journal of Hospitality Management, 2012, 31(1): 191-198.

[23] Liu S, Law R, Rong J, et al. Analyzing Changes in Hotel Customers' Expectations by Trip Mode [J]. International Journal of Hospitality Management, 2013, 34(1): 359-371.

[24] 保继刚. 旅游者行为研究[J]. 社会科学家, 1987(6): 19-22. (Bao Jigang. Research the Travelers' Behavior[J]. Social Scientist, 1987(6): 19-22.)

[25] 杨瑞. 西安市大学生旅游行为模式研究[D]. 西安: 陕西师范大学, 2006. (Yang Rui. The Study on the University Student's Travel Behavior Pattern in Xi'an [D]. Xi'an: Shaanxi Normal University, 2006.)

[26] 岳冬菊, 杨媛. 西安市国内游客旅游偏好实证分析[J]. 西安文理学院学报: 自然科学版, 2010, 13(2): 116-119. (Yue Dongju, Yang Yuan. An Empirical Analysis of Domestic Tourists' Traveling Preferences in Xi'an [J]. Journal of Xi'an University of Arts & Science: Natural Science Edition, 2010, 13(2): 116-119.)

[27] 雷俐丽. 大连国内游客旅游偏好及行为特征研究[D]. 大连: 东北财经大学, 2011. (Lei Lili. Research on Tourism Preference and Behavior Characteristics of Domestic Tourist in Dalian [D]. Dalian: Dongbei University of Finance & Economics, 2011.)

[28] 顾秀玲. 环太湖度假酒店四类顾客消费行为实证研究[J].

牡丹江师范学院学报: 哲学社会科学版, 2014(5): 35-38. (Gu Xiuling. An Empirical Research of Consumers' Behavior [J]. Journal of Mudanjiang Normal University: Philosophy Social Sciences Edition, 2014(5): 35-38.)

[29] Zhao W X, Jiang J, Yan H, et al. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid [C]. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. 2010: 56-65.

[30] Titov I, McDonald R. Modeling Online Reviews with Multi-grain Topic Models [C]. In: Proceedings of the 17th International Conference on World Wide Web. ACM, 2008: 111-120.

[31] Kasper W, Vela M. Sentiment Analysis for Hotel Reviews[C]. In: Proceedings of Computational Linguistics Applications Conference. 2012.

[32] Chen Y S, Chen L H, Takama Y. Proposal of LDA-Based Sentiment Visualization of Hotel Reviews [C]. In: Proceedings of the 2015 IEEE International Conference on Data Mining. 2015: 687-693.

[33] Blei D M. Probabilistic Topic Models [J]. Communications of the ACM, 2012, 55(4): 77-84.

[34] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[35] Yoo K H, Gretzel U. What Motivates Consumers to Write Online Travel Reviews? [J]. Information Technology & Tourism, 2008, 10(4): 283-295.

[36] Duan H, Zirn C. Can We Identify Manipulative Behavior and the Corresponding Suspects on Review Websites Using Supervised Learning? [A]. // Secure IT Systems [M]. Springer Berlin Heidelberg, 2012: 215-230.

[37] Wu G, Greene D, Cunningham P. Merging Multiple Criteria to Identify Suspicious Reviews [C]. In: Proceedings of the 4th ACM Conference on Recommender Systems. ACM, 2010: 241-244.

[38] Blanke J, Chiesa T. The Travel & Tourism Competitiveness Index 2011: Assessing Industry Drivers in the Wake of the Crisis [R]. World Economic Forum, 2011.

作者贡献声明:

杨海霞: 提出研究思路, 设计研究方案, 进行实验, 论文起草及最终版本修订;
吴维芳: 修改论文;
孙含林: 采集、清洗和分析数据, 修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: haixiayang@whu.edu.cn。
[1] 杨海霞, 吴维芳, 孙含林. LasVegas.csv. 拉斯维加斯酒店评论数据.

收稿日期: 2016-06-12
收修改稿日期: 2016-07-16

Analyzing Travelers' Preferences for Hotels Based on Structural Topic Model

Yang Haixia Wu Weifang Sun Hanlin
(Economics and Management School, Wuhan University, Wuhan 430072, China)

Abstract: [**Objective**] This paper aims to identify various types of travelers' preferences for hotel services. [**Methods**] First, we classified the hotels as luxury and budget ones, and then divided the travelers into five categories. Second, we analyzed individual traveler's rating behaviors on the hotel review website TripAdvisor. Finally, we analyzed the latent topics of hotel reviews with the help of Structure Topic Model (STM) to identify travellers' preferences for hotel services. [**Results**] We found that the average rating scores of luxury hotels were higher than the budget ones and travelers did have different preferences for hotel services. [**Limitations**] The dataset for our study was not large enough. We did not consider the impacts of gender and age to hotel rating and online review contents. [**Conclusions**] Analyzing travelers' preferences for hotels could help both the managers and travelers make right decisions.

Keywords: Online review Patterns of rating Hotel grade Travelers' profiles STM Review topics